

DOCUMENT RESUME

ED 171 751

TR 000 952

AUTHOR Beard, Jacob G.; Pettie, Allan L.
TITLE A Comparison of Linear and Rasch Equating Results for Basic Skills Assessment Tests.
PUB DATE Apr 79
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basic Skills; Communication Skills; Comparative Statistics; Complexity Level; Elementary Education; Elementary School Mathematics; *Equated Scores; *Goodness of Fit; Grade 3; Grade 5; *Item Analysis; *Mastery Tests; *Test Items
IDENTIFIERS Florida Statewide Assessment Program; *Linear Equating Method; *Rasch Scaled Scores

ABSTRACT

Test results from the Florida Educational Assessment of third and fifth grade communications and mathematics skills were used to compare linear and Rasch equating results. The samples consisted of over 5,000 cases for each grade and content area. The tests contained some items common to both the 1976 and 1977 test forms, but no fewer than 20 percent of the total number of items in each test. Item discrimination and shifts in calibration were examined. Analysis indicated that there was little instability in each set of common items; that each set of common items contained a low proportion of misfitting items based on the discrimination index; and that the total tests contained a low proportion of misfitting items compared to Anchor Test Study data. Results indicated close similarity in reference T scores and high comparability between linear and Rasch equating results. Rasch procedures appeared to be applicable to negatively skewed mastery tests of basic skills. The common items demonstrated stability in difficulty parameter estimates, and very few items did not fit the model using the discrimination criterion. It was concluded that Rasch equating was comparable to linear equating in the analysis of longitudinal trends in basic skills achievement. (MH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

D-43

A Comparison of Linear and Rasch
Equating Results for Basic Skills
Assessment Tests¹

Jacob G. Beard and Allan L. Pettie²

Florida State University

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Jacob G. Beard

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

The purpose of this study was to determine if Rasch equating procedures are applicable to negatively skewed mastery tests of basic skills. Theoretical considerations are discussed in the first section of the report. The second section deals with anchor tests and the analysis of common items with Rasch scaling techniques. Finally, a comparison is drawn between linear and Rasch equating using common items.

Under the existing format of the Florida Educational Assessment, different forms of third and fifth grade communications and mathematics basic skills tests are administered each year. However, each of the tests contained items common to both the 1976 and 1977 forms. The common items were identified for each pair of tests. These common items were then used to provide adjustments to the mean and variance in the linear equating method and to provide additive constants which were necessary to adjust the ability scale of tests in the Rasch equating procedure. The data were used to compare linear and Rasch equating results.

The data base for this study consisted of 5 percent samples drawn systematically from the assessment files for each grade and each content area. The systematic samples consisted of more than five thousand cases for each grade and content area.

¹Paper presented at the 1979 meeting of the American Educational Research Association, San Francisco, California.

²The assistance of Leslie J. Caldwell is gratefully acknowledged.

ED171751

TM008 952

THEORETICAL CONSIDERATIONS

Angoff(1971) presents the details of linear and other forms of equating. According to him, "Equivalent scores" can be defined as scores which correspond to the same ability. This definition is analogous to the conventional definition of equivalent scores - "Two scores. . . may be considered equivalent if their corresponding percentile ranks in any given group are equal". (Angoff, W. H. In Thorndike, R. L., Educational Measurement, p. 563) Rasch equating defines equivalent scores as scores which correspond to the same Rasch "log" ability.

Prior to raw score-to-raw score equating, calibration of each test onto the same scale must be performed. It is necessary to make a distinction between equating and calibration. According to Angoff(1971,p. 565), calibration denotes the specification of abilities to raw scores, but equating denotes the determination of equivalent raw scores. Linear equating is a method of score calibration in which the ability level and range spanned by the test are reflected on the reference scale scores. The details of equating with common items given to two separate nonrandom groups on two different tests are provided by Angoff(1971). Briefly, the major assumptions of this form of linear equating are that the regression systems for the two groups of subjects would have been identical if the two groups had taken the same test, that the common item set represents the same psychological function in both groups, and that the two groups do not differ very much in ability (the groups are not assumed to be equivalent).

Rasch model equating is defined as the assignment of scores which correspond to the same log ability. George Rasch(1973) proposed the construction of tests with a specific frame of reference. One important

principle of the Rasch model has been labeled "specific objectivity". Rasch(1966) proposed the creation of measurements such that person comparisons in the target population or task comparisons in the domain of tasks are invariant with respect to the specific sample of tasks or people. "Specific objectivity" refers to this invariance property. Specific objectivity has been more formally defined by Rasch as". . . whenever the comparison of any two parameters within the same set may be carried out in such a way that it (the comparison) is unaffected by all other unknown parameters in the system. . . the comparison is characterized as 'specifically objective'." The two parameters referred to are - one for persons (ability) and the other for items (difficulty). For example, the determination of ability parameters is not dependent upon the difficulty parameters of the items.

The comparison of linear and Rasch equating with common items was accomplished by the use of the same data. The four tests used were mathematics and communications skills at the third and fifth grade levels. The tests were calibrated separately. These calibrations resulted in a separate ability table for each test at each grade level.

Wright(1977) presented a method for equating two tests by the use of common items. This procedure does not require the administration of two tests to the same group. Rentz and Bashaw(1975,1977) describe two methods for accomplishing the equating task. In these methods a reference scale is derived for linking items for separate tests. This reference scale is normally the ability scale (expressed in natural log units) derived from a Rasch test analysis. If one has two tests which are analyzed separately, then the ability scale defined for each test has the same units with an arbitrary origin. This arbitrary origin is normally determined by setting the average item difficulty to zero, which coin-

4

cidently fixes the origin of the ability scale. Since there are two tests involved, it is necessary to estimate the difference between the averages of item difficulty for the two tests. This estimated difference in item difficulty averages is a necessary adjustment (additive constant) to place both item difficulty and ability values onto a scale with a common origin.

To estimate this additive constant requires the implementation of one of two methods. Method one requires the administration of two tests to one group of people. Method two requires the administration of two tests to two different samples of people. The first method would be an appropriate strategy to scale alternate test forms onto a common scale. The second method is an appropriate strategy to link together tests of different difficulty. Because two samples of people were administered two different tests and because the tests were noted to be of different difficulty, method two was the strategy used in this study.

Some further elaboration of the procedures necessary for Method two is required. The principle of this strategy is that the common item set should have the same average item difficulty for both tests. Any difference in average item difficulty of the common items between the two tests indicates that the common items were more difficult than non-common items in one test and easier than non-common items in the other test. This difference in average item difficulty for the core items can be used as an adjustment factor to adjust the scale of either of the tests to the scale of the other test.

Wright(1977) presented a convenient formula to determine the constant necessary to translate all item difficulties in test 2 calibration onto the scale of base test 1. This formula is:

$$T_{12} = \frac{\sum_{i=1}^k (D_{i1} - D_{i2})}{K}$$

where T_{12} equals the additive constant
 K equals the number of common items
 D_{i1} equals the item difficulties on test 1
 D_{i2} equals the item difficulties on test 2

The standard error for the additive constant is about equal to $3.5 / \sqrt{NK}$ logits (Wright, 1977). The divisor N is equal to the number of people in the sample. Logits are the units of the Rasch scale.

DESCRIPTION AND ANALYSIS OF COMMON ITEMS

Wright (1977) provides the general guidelines that a sample of 400 people and 10 good items have been ascertained to be adequate to maintain link quality between two tests. For this study the smallest number of common items was 20. Since the original equating task followed the Design IV linear equating procedures described by Angoff (1971), the common item sets contained no fewer than 20 items or 20% of the total number of items in each test. An examination of the validity of the common item set was given in the original equating report (Beard, Caldwell, Green, and Pettie, 1978).

The validity of the common item set was reexamined using Rasch analysis. The following methods were used: 1) the discrimination index for each item was calculated and evaluated, 2) graphic presentation procedures developed by Rentz (1978), were used to detect shifts in the item

calibration from 1976 to 1977. Discrimination values should be near unity for fitting items. This discrimination index or slope is the regression of the "item log odds" on "test log odds". A discrimination value of less than one indicates that the item characteristic curve for that item is flatter than the test characteristic curve. Conversely, a slope value greater than one indicates that the item characteristic curve for that item is steeper than the test characteristic curve. An overall measure of the fit of Rasch model equating analysis is called test fit. Two different aspects of test fit were examined. The first aspect of test fit can be considered an antecedent condition of the model. This aspect is the extent to which the test contains items which fit the model. The second aspect can be defined as the stability of the Rasch difficulty parameter estimates. Both these aspects of test fit will be discussed in the next section.

Figure 1 illustrates the 1976 Third Grade Communications common item difficulty values plotted against those of 1977. The difficulty estimates represented in plots of this kind should be located small distances from a unit slope straight line. The intercept of this unit slope line is equal to the mean difference (additive constant) between the two calibrations. By examining Figure 1 it can be seen that the Third Grade Communications common items appear to fulfill the prediction of the model. The three items which fall above and relatively far from the unit slope line had calibration discrimination values which fall outside the recommended range of 1 plus or minus 0.2 (Rentz and Bashaw, 1975). These items would not have been retained under item fit analysis; however, these calibrations occurred after the second testing and were included in the 1977 test, and in the linear equating. Thus, it would be improper to

exclude the items at this juncture. The 1976 Third Grade Communications test had 58 items out of 81 total items and 26 items out of 29 common items having discrimination indices in the range of 1 plus or minus .2. The 1977 Third Grade Communications test had 63 items out of 87 total items and 22 out of 29 common items fit the model. From these results it was concluded that Rasch equating techniques are appropriate for the Third Grade Communications test.

FIGURE 1

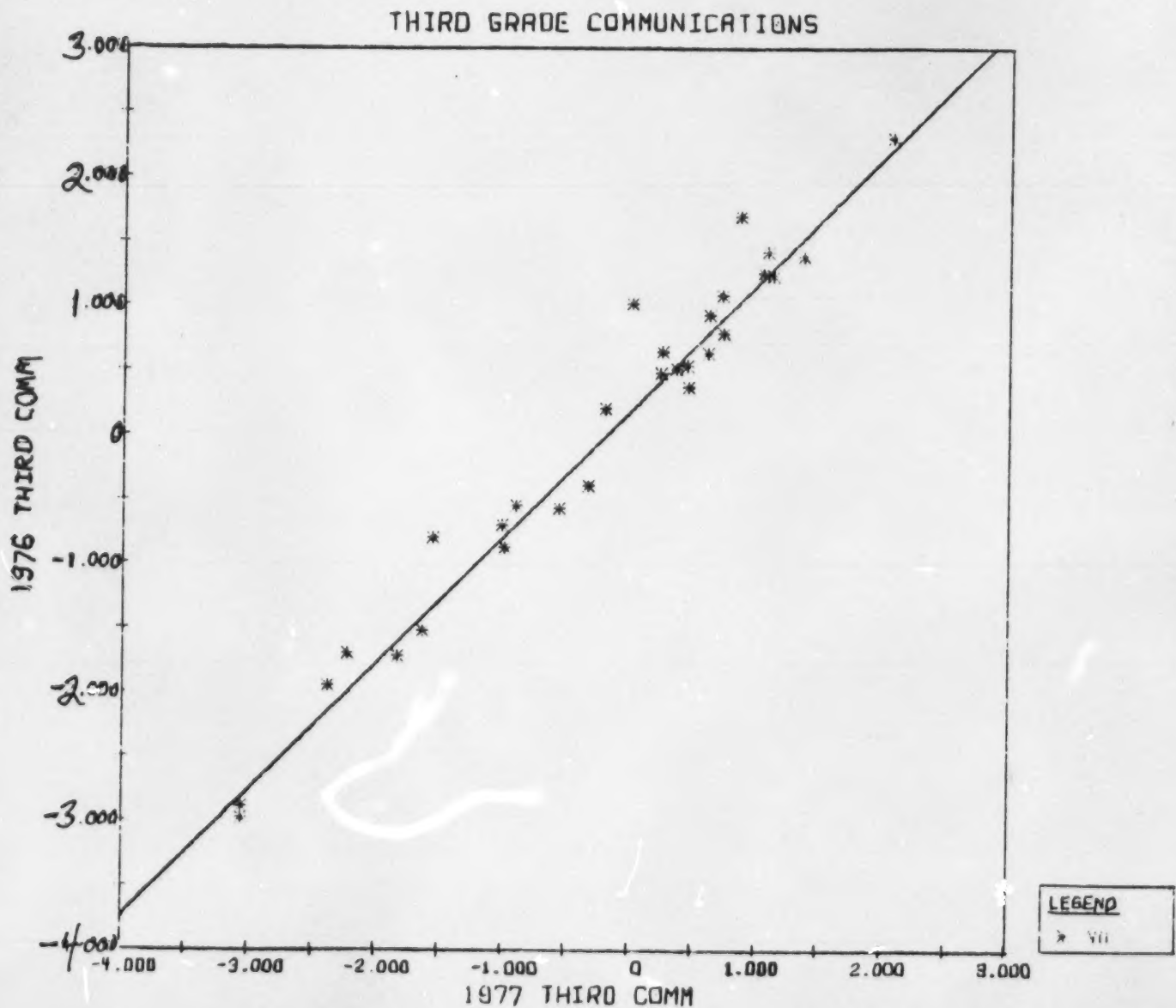


Figure 2 demonstrates another example of proper behavior of the common item set according to the prediction of the model. The item difficulties represent a range of about -2.0 to 2.0 . This spread in item difficulty provides a sound basis for calibration. Using the discrimination criterion, the 1976 Third Grade Mathematics test had 51 items out of 64 total items and 16 out of 21 common items fit the model. The 1977 Third Grade Mathematics test had 51 items out of 59 total items and 19 out of 21 common items fit the model. The Third Grade Mathematics tests were appropriate to Rasch equating analysis.

Figure 3 shows that for Fifth Grade Communications, fit to the unit slope line is good, and there is little reason to suspect that the common items for the Fifth Grade Communications test have difficulty parameter estimates which were poorly estimated. Again, the item difficulty estimates vary in range from -2.0 to 2.0 . Using the discrimination index criterion, the 1976 Fifth Grade Communications test had 126 items out of 136 total items and 30 out of 34 common items fitting the model. The 1977 Fifth Grade Communications test had 121 items out of 141 total items and 29 out of 34 common items fitting the model. From the plot and the analysis of the discrimination values, there is little evidence to indicate that the Fifth Grade Communications tests were not compatible with Rasch equating analysis.

Although the common item difficulty values for Fifth Grade Mathematics in Figure 4 have a small range, the fit of the items looks good. Two of the three poor fitting items had 1976 discrimination values below 0.8 . The other poor fitting item had a discrimination value below 0.8 for 1977. Otherwise, the common items appear to have the consequent condition

FIGURE 2

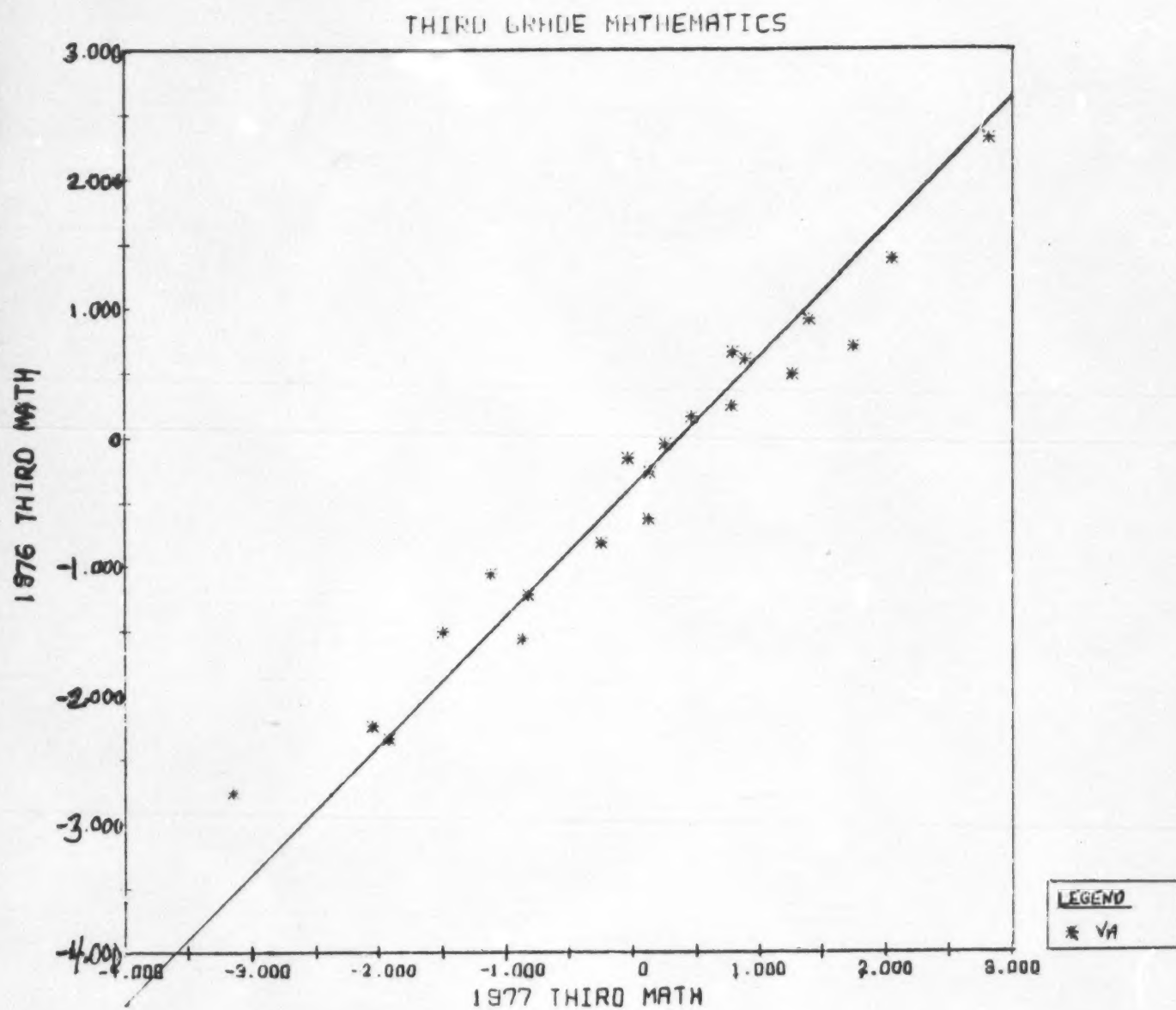


FIGURE 3

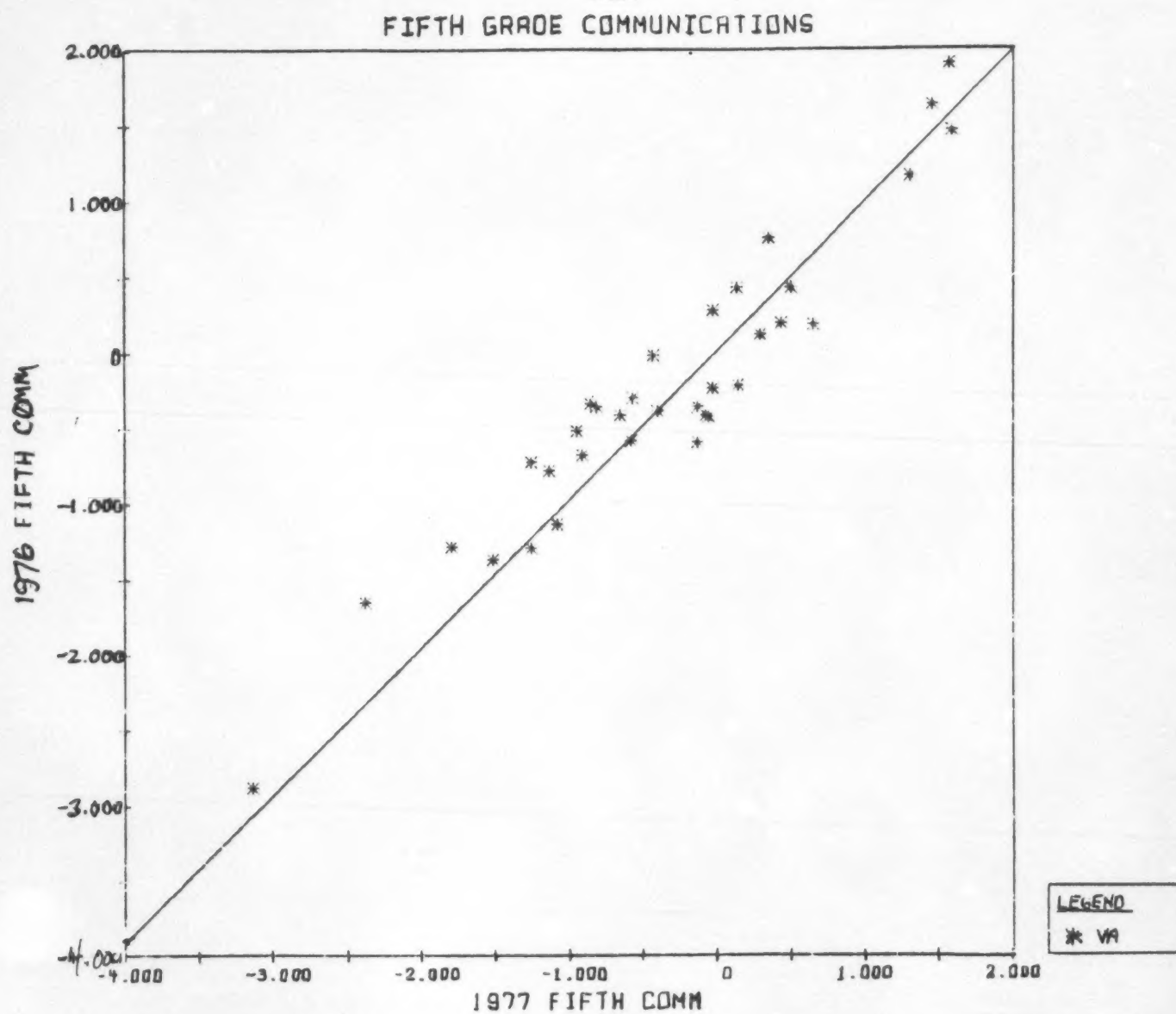
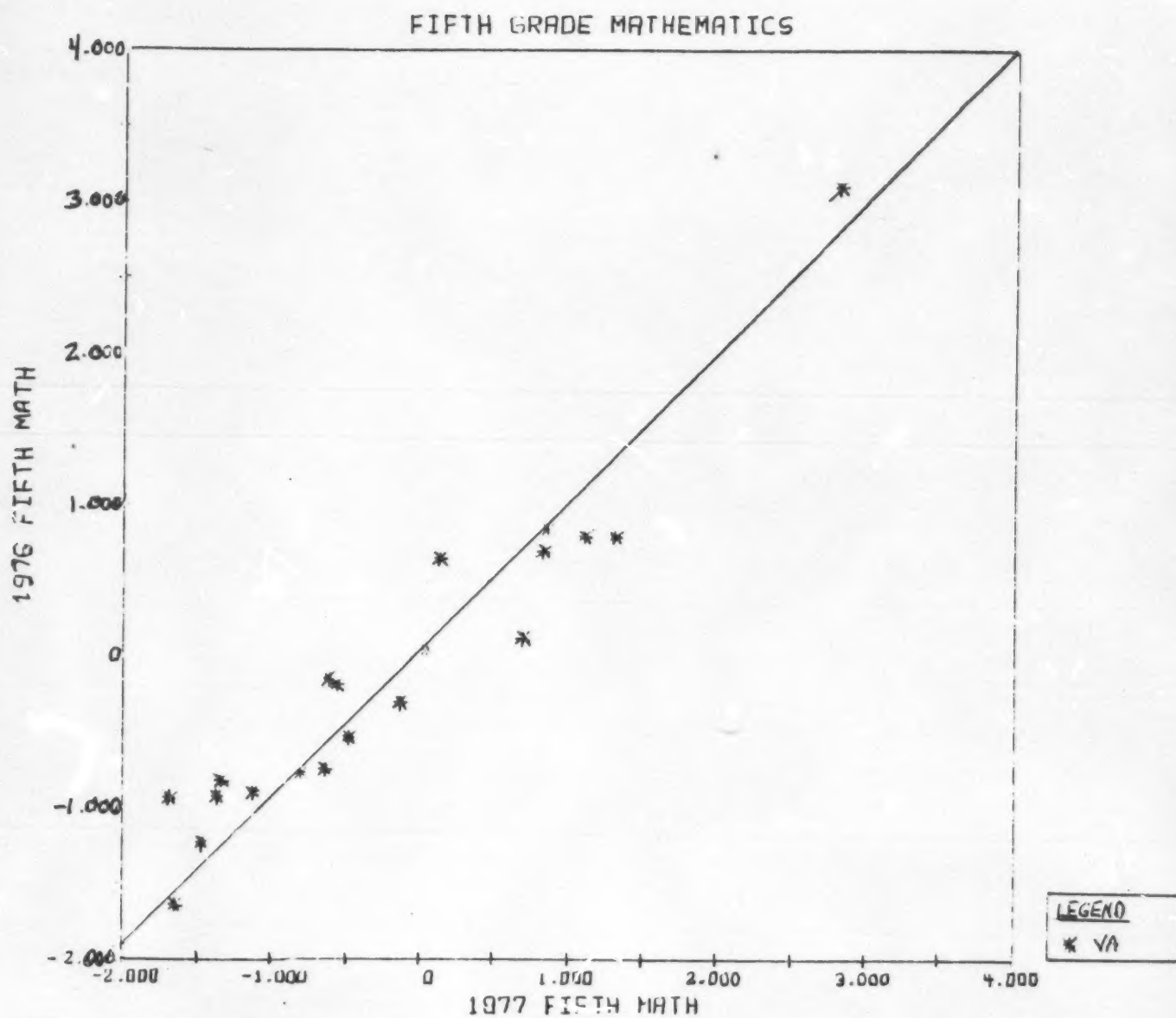


FIGURE 4



of stability. The discrimination values for the 1976 Fifth Grade Mathematics items showed good overall fit. Only 7 items out of 74 total items fell outside the discrimination confidence interval. For the common items only 4 items in 1976 and 5 items in 1977 out of 20 common items failed to fit the model.

From the above analysis it can be seen that there is little instability in each set of common items, that each set of common items contained a low proportion of misfitting items based on the discrimination index, and that the total tests contained a low proportion of misfitting items compared to the Anchor Test Study data (Rentz and Bashaw, p. 42), in which approximately 45 percent of the items were poor fits to the model. In a preceding section test fit was defined as the extent to which the test contained fitting items and as a consequent condition of specific objectivity. This specific consequent condition concerns the stability of each difficulty parameters. It has been shown from the plot that these estimates are stable.

RESULTS

Table 1 presents the statistical characteristics of the sample data. These sample statistics were compared to population values and were found to be almost equivalent.

TABLE 1
Statistical Characteristics of the Sample

Grade	Subtest	Year	Mean	Mean P Value	Number of Items	Variance	Standard Deviation
3	Communications	1976	67.031	.828	81	152.940	12.367
3	Communications	1977	73.777	.848	87	189.520	13.767
3	Mathematics	1976	53.806	.841	64	98.247	9.912
3	Mathematics	1977	52.991	.898	59	50.081	7.077
5	Communications	1976	114.663	.843	136	472.206	21.730
5	Communications	1977	120.091	.852	141	485.100	22.024
5	Mathematics	1976	55.563	.751	74	187.772	13.703
5	Mathematics	1977	74.358	.800	93	252.771	15.899

Table 2 presents the statistical characteristics of the common item set. It can be seen that the mean scores, and consequently the mean p values, are higher in 1977 than 1976 for both grade levels and both content areas. In the analysis of the common items, an average increase of from 1.3 percentage point mean change to 4.2 percentage point mean change was observed.

TABLE 2
Statistical Characteristics of the Anchor Tests

Grade	Subtest	Year	Mean	Mean P Value	Number of Items	Variance	Standard Deviation
3	Communications	1976	24.024	.828	29	22.788	4.774
3	Communications	1977	25.022	.863	29	17.102	4.136
3	Mathematics	1976	18.235	.868	21	10.212	3.196
3	Mathematics	1977	18.702	.891	21	7.385	2.718
5	Communications	1976	29.652	.872	34	31.149	5.581
5	Communications	1977	30.085	.885	34	28.034	5.295
5	Mathematics	1976	15.514	.776	20	14.427	3.798
5	Mathematics	1977	16.363	.818	20	11.247	3.354

Table 3 presents the mean Rasch difficulties for the common item sets. It can be seen that for the Third and Fifth Grade Communications and Fifth Grade Mathematics that the common items were easier than the non-common items for the 1977 test relative to the 1976 non-common items. In contrast, the common item set for the Third Grade Mathematics test were easier than the non-common items for the 1976 test and harder than the non-common items for the 1977 test. This table also contains the additive constants which were used to adjust the scale of the 1976 test to the scale of the 1977 reference scale. These additive constants were derived by subtracting the 1976 mean difficulty value from the 1977 base test mean difficulty value. For example, the additive constant for Third Grade Communications (-.261) was obtained by subtracting -.006 (the 1976 value) from -.267 (the 1977 value). The final component of this table is the standard error of each additive constant. These standard errors are small in value. Appendix A contains the difficulty values for the common items by content area and grade.

TABLE 3
Rasch Difficulties Estimates for the Common Items

Grade	Subtest	Year	Number of Items	Mean Difficulty	Standard Deviation	Additive Constant	Standard Error
3	Communications	1976	29	-.006	1.369	-.261	.009
3	Communications	1977	29	-.267	1.359		
3	Mathematics	1976	21	-.346	1.301	.398	.010
3	Mathematics	1977	21	+.052	1.479		
5	Communications	1976	34	-.246	.957	-.093	.008
5	Communications	1977	34	-.339	1.059		
5	Mathematics	1976	20	-.102	1.057	-.099	.011
5	Mathematics	1977	20	-.201	1.182		

In the original linear equating report the 1977 distributions were chosen as the reference scales. The results were reported as non-normalized T scores. The T scores reported in Table 4 were derived by: 1) converting the 1976 raw score to its corresponding ability scale value, 2) applying the additive constant to this 1976 ability value, and 3) locating the 1977 ability value which is nearest in value to the adjusted 1976 ability value. The 1977 ability value derived by this process was then converted back to a raw score on the 1977 scale. Each raw score was then transformed to the 1977 T score which is in Table 4. It is important to note that the largest discrepancy between the linear equating and Rasch equating is 6 units on the T scale. This discrepancy occurs at the lower end of the distribution where there was little data. There are very small differences between the two equating techniques. Finally, Rasch equated values are smaller than corresponding linear equated values generally throughout the range of 1976 raw scores.

The results presented in Table 5 parallel the results for Table 4. These T scores were derived by the same procedures as presented above. Table 4 presents results for Third Grade Communications and Mathematics. Table 5 presents results for Fifth Grade Communications and Mathematics.

TABLE 4

Comparison of Linear and Rasch Equating For 1976 Third Grade
Raw Scores Placed Onto the Reference Scale

Third Grade Communications				Third Grade Mathematics			
1976 Score				1976 Score			
Raw	T	Equated T		Raw	T	Equated T	
		Linear	Rasch			Linear	Rasch
81	61			64	60		
80	60	60	59	63	59	59	57
79	60	59	57	62	58	58	56
78	59	58	57	61	57	57	56
77	58	57	56	60	56	56	54
76	57	56	55	59	55	54	53
75	56	55	54	58	54	53	53
74	56	54	52	57	53	52	51
73	55	53	52	56	52	51	50
72	54	52	51	55	51	50	49
71	53	51	50	54	50	49	48
70	52	51	50	53	49	47	47
69	52	50	49	52	48	46	46
68	51	49	49	51	47	45	44
67	50	48	47	50	46	44	44
66	49	47	46	49	45	43	43
65	48	46	46	48	44	42	42
64	48	45	45	47	43	41	40
63	47	44	44	46	42	39	39
62	46	43	44	45	41	38	37
61	45	42	42	44	40	37	37
60	44	41	40	43	39	36	35
59	44	40	40	42	38	35	34
58	43	40	40	41	37	34	33
57	42	39	39	40	36	32	32
56	41	38	39	39	35	31	32
55	40	37	37	38	34	30	30
54	39	36	36	37	33	29	28
53	38	35	36	36	32	28	27
52	38	34	35	35	31	27	26
51	37	33	34	34	30	25	25
50	36	32	33	33	29	24	23
49	35	31	32	32	28	23	22
48	35	30	31	31	27	22	22
47	34	30	31	30	26	21	20
46	33	29	30	29	25	20	19
45	32	28	29	28	24	19	18
44	31	27	28	27	23	17	16
43	31	26	28	26	22	16	15
42	30	25	27	25	21	15	13
41	29	24	27	24	20	14	12
40	28	23	25	23	19	13	10
39	27	22	25	22	18	12	10
38	27	21	24	21	17	10	8
37	26	20	23	20	16	9	6
36	25	19	23	19	15	8	5
35	24	19	20	18	14	7	3
34	23	18	20	17	13	6	2
33	22	17	19	16	12	5	
32	22	16	18	15	11	3	1
31	21	15	17	14	10	2	0
30	20	14	17	13	9	1	
29	19	13	16	12	8	0	
28	18	21	15				
27	18	11	14				
26	17	10	13				
25	16	9	12				
24	15	9	12				
23	14	8	11				
22	14	7	10				
21	13	6	9				
20	12	5	9				
19	11	4	8				
18	10	3	7				
17	10	2	7				
16	9	1	6				
15	8	0	5				

TABLE 5

Comparison of Linear and Rasch Equating For 1976 Fifth Grade
Raw Scores Placed Onto the Reference Scale

Fifth Grade Communications				Fifth Grade Mathematics			
1976 Score				1976 Score			
Raw	T	Equated T		Raw	T	Equated T	
		Linear	Rasch			Linear	Rasch
134-135	59	59	59	73	63	62	61
133	58	58	58	72	62	61	60
132	58	58	57	71	61	60	59
131	58	57	57	70	61	59	58
130	57	57	56	69	60	59	58
129	57	56	56	68	59	58	57
128	56	56	55	67	58	57	56
127	56	55	54	66	58	56	55
126	55	55	54	65	57	55	54
125	55	54	53	64	56	55	54
124	54	54	53	63	55	54	53
123	54	53	52	62	55	53	52
122	53	53	52	61	54	52	51
121	53	52	51	60	53	51	51
120	52	52	51	59	53	51	51
119	52	51	50	58	52	50	49
118	52	51	50	57	51	49	49
117	51	50	49	56	50	48	48
116	51	50	49	55	50	47	47
115	50	49	48	54	49	46	46
114	50	48	47	53	48	46	46
113	49	48	47	52	47	45	45
112	49	47	46	51	47	44	44
111	48	47	46	50	46	43	43
110	48	46	45	49	45	42	42
109	47	46	45	48	44	42	42
108	47	46	45	47	44	41	40
107	46	45	44	46	43	40	40
106	45	45	44	45	42	39	38
105	45	44	43	44	42	38	38
104	44	44	43	43	41	38	38
103	44	43	42	42	40	37	37
102	43	43	42	41	39	36	35
101	43	42	41	40	39	35	35
100	42	42	41	39	38	34	34
99	42	41	40	38	37	34	34
98	42	41	40	37	36	33	33
97	41	40	40	36	36	32	32
96	41	40	39	35	35	31	31
95	41	40	39	34	34	30	29
94	40	39	38	33	34	29	29
93	40	38	37	32	33	29	28
92	39	38	37	31	32	28	27
91	39	37	36	30	31	27	26
90	38	37	36	29	31	26	25
89	38	36	35	28	30	25	24
88	37	36	35	27	29	25	23
87	37	35	34	26	28	24	23
86	36	35	34	25	28	23	22
85	36	34	33	24	27	22	21
84	35	34	33	23	26	21	20
83	35	33	32	22	26	21	20
82	35	33	32	21	25	20	19
81	34	33	32	20	24	19	18
80	34	32	31	19	23	18	17
79	33	32	31	18	23	17	16
78	33	31	30	17	22	16	15
77	32	31	30	16	21	16	15
76	32	30	30	15	20	15	14
75	31	30	29	14	20	14	13
74	31	29	29	13	19	13	12
73	30	29	28	12	18	12	11
72	30	28	27	11	17	12	11
71	29	28	27	10	17	11	10
70	29	27	26	9	16	10	9
69	28	27	26	8	15	9	8
68	28	26	25	7	15	8	7
67	27	26	25	6	14	8	7
66	27	25	25	5	13	7	6
65	26	25	24	4	12	6	6
64	26	24	24	3	12	5	5
63	25	24	24	2	11	4	4
62	25	24	24				

TABLE 5 (CONT'D)

Fifth Grade Communications			
1976 Score			
Raw	T	Equated T	
		Linear	Rasch
61	25	23	23
60	25	23	23
59	24	22	22
58	24	22	22
57	23	21	21
56	23	21	21
55	23	20	20
54	22	20	20
53	22	19	20
52	21	19	19
51	21	19	18
50	20	18	18
49	20	18	18
48	19	17	17
47	19	17	16
46	18	16	16
45	18	16	15
44	17	15	15
43	17	15	15
42	17	14	15
41	16	14	14
40	16	13	14
39	15	13	13
38	15	12	12
37	14	12	12
36	14	11	11
35	13	11	11
34	13	10	10
33	12	10	10
32	12	9	9
31	11	9	9
30	11	8	8
29	11	8	8
28	10	7	7
27	10	7	7
26	9	6	6
25	9	6	5
24	8	5	4
23	8	5	4
22	7	5	4
21	7	4	3
20	6	4	3
19	6	3	3
18	6	3	2
17	5	2	2
16	5	2	2
15	4	1	1
14	4	1	1

The T scores reported in Table 6 were derived by converting the mean raw scores to T scores on the reference scale. Because the 1977 scores were designated as the reference scale, their mean raw scores convert to a score of 50 on the reference scale. The 1976 mean reference scores are all less than 50 on the reference T scale. Table 6 also serves as convenient means to compare linear and Rasch equating. The 1976 mean reference scores by both methods are remarkably close, but the Rasch values tend to be slightly smaller than the linear equating results. This remarkable similarity in reference T scores denotes high comparability between the two methods.

TABLE 6
Raw Score Means Converted to the Reference Scale
by Rasch and Linear Equating

Grade	Subtest	Year	Linear T Score	Rasch T Score
3	Communications	1976	47.81	47.58
3	Communications	1977	50.00	50.00
5	Communications	1976	49.23	49.05
5	Communications	1977	50.00	50.00
3	Mathematics	1976	48.43	48.06
3	Mathematics	1977	50.00	50.00
5	Mathematics	1976	47.72	46.54
5	Mathematics	1977	50.00	50.00

From the results of this report differences between the linear and the Rasch equating were so small as to appear inconsequential. Rasch procedures appear to be applicable to negatively skewed mastery tests of basic skills. The common items demonstrated stability in difficulty parameter estimates and very few of the items did not fit the model using the discrimination criterion. The total test fit was better than the tests analyzed by Rentz and Bashaw(1975). Therefore, it was concluded that Rasch equating is comparable to linear equating in the analysis of longitudinal trends in student achievement of the basic skills objectives.

REFERENCES

- Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement. (2nd ed.). Washington, D. C.: American Council on Education, 1971.
- Beard, J. G., Caldwell, L. J., Green, J. C., and Pettie, A. L. Equating Florida's 1976 and 1977 basic skills assessment scores: Report of method and results for third and fifth grades. Tallahassee, FL: Florida State University, Educational Research and Evaluation Services, 1978.
- Rasch, G. An item analysis that takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19(1), 49-57 (b).
- Rentz, C. C. Monitoring the quality of an item-pool calibrated by the Rasch model. Paper presented to National Council in Measurement in Education, Toronto, Ontario, March, 1978.
- Rentz, C. C. and Bashaw, W. L. Equating Reading Tests with the Rasch Model, Volume I, Final Report. Athens, GA: University of Georgia, Educational Research Laboratory, 1975. (ERIC Document Reproduction Nos. ED 127 330).
- Rentz, C. C. and Bashaw, W. L. The national reference scale for reading: An application of the Rasch Model. Journal of Educational Measurement, 1977, 14, 161-169.
- Wright, B. D. The Rasch Model. Journal of Educational Measurement, 1977, 14, 97-116.